

Exploring the Impact of 3D Fast Spin Echo and Inversion Recovery Gradient Echo Sequences Magnetic Resonance Imaging Acquisition on Automated Brain Tumor Segmentation

Mana Moassefi, MD; Shahriar Faghani, MD; Sara Khanipour Roshan, MD; Gian Marco Conte, MD, PhD; Seyed Moein Rassoulinejad Mousavi, MD; Timothy J. Kaufmann, MD; and Bradley J. Erickson, MD, PhD

Abstract

Objective: To conduct a study comparing the performance of automated segmentation techniques using 2 different contrast-enhanced T1-weighted (CET1) magnetic resonance imaging (MRI) acquisition protocol.

Patients and Methods: We collected 100 preoperative glioblastoma (GBM) MRIs consisting of 50 IR-GRE and 50 3-dimensional fast spin echo (3D-FSE) image sets. Their gold-standard tumor segmentation mask was created based on the expert opinion of a neuroradiologist. Cases were randomly divided into training and test sets. We used the no new UNet (nnUNet) architecture pretrained on the 501-image public data set containing IR-GRE sequence image sets, followed by 2 training rounds with the IR-GRE and 3D-FSE images, respectively. For each patient, in the IR-GRE and 3D-FSE test sets, we had 2 prediction masks, one from the model fine-tuned with the IR-GRE training set and one with 3D-FSE. The dice similarity coefficients (DSCs) of the 2 sets of results for each case in the test sets were compared using the Wilcoxon tests.

Results: Models trained on 3D-FSE images outperformed IR-GRE models in lesion segmentation, with mean DSC differences of 0.057 and 0.022 in the respective test sets. For the 3D-FSE and IR-GRE test sets, the calculated *P* values comparing DSCs from 2 models were .02 and .61, respectively.

Conclusion: Including 3D-FSE MRI in the training data set improves segmentation performance when segmenting 3D-FSE images.

© 2024 THE AUTHORS. Published by Elsevier Inc on behalf of Mayo Foundation for Medical Education and Research. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) ■ Mayo Clin Proc Digital Health 2024;2(2):231-240

Glioblastoma multiforme (GBM), the most common primary malignant brain tumor in adults, accounts for 77%-81% of all primary malignant tumors of the central nervous system.¹ More than 10,000 Americans could have been diagnosed with GBM in 2023 based on a 3.19-per-100,000 incidence rate.² Magnetic resonance imaging (MRI) provides detailed brain structure for diagnosing GBMs. On MRI, GBM tumors are located and semantically segmented, and the tumor size is measured. The findings derived from this process enable radiologists and clinicians to determine the tumor components, size, and aggressiveness and

plan further treatment, such as surgical and radiation therapy.³

In contrast to some tumors with more homogeneous structures, like meningiomas, GBMs are diffusely infiltrative, heterogeneously enhancing, and frequently centrally necrotic and have asymmetrical shapes and extended tentacle-like structures, making them more complex and controversial to delineate.⁴ Moreover, unlike computed tomography scans, MRI does not possess a standardized or absolute voxel value scale to allow thresholding. Variations in scanners (field strength of 1.5, 3, or 7 T) and image acquisition protocols (eg, field of view, voxel resolution, gradient strength,

From the Mayo Clinic Artificial Intelligence Laboratory, Department of Radiology (M.M., S.F., G.M.C., S.M.R.M., B.J.E.), and Department of Radiology (M.M., S.F., S.K.R., G.M.C., S.M.R.M., T.J.K., B.J.E.), Mayo Clinic, Rochester, MN.

and several other parameters such as repetition and echo times [TE]) can result in substantial differences in grayscale values of the same tumor across different imaging centers. These factors complicate automatic tumor segmentation. Therefore, expertise and considerable time and cost are necessary for manual GBM segmentation on MRI.

Automated GBM segmentation using deep learning models offers a cost-efficient solution that significantly reduces the time required for this time-consuming task.⁵ This approach can be implemented with or without the further supervision of radiologists, providing a consistent and efficient workflow.⁶ By using the same algorithm for all cases, the segmentation process becomes more consistent compared with the variability introduced by multiple radiologists' opinions. Accurate segmentation plays a crucial role in radiation planning, response assessment across multiple examinations, and the evaluation of specific tumor regions using advanced imaging techniques such as perfusion and diffusion parameters within contrast-enhancing portions of the tumor.⁷ Performing GBM tumor segmentation using automated deep learning-based tools has shown great promise in recent years.^{8–12} In 2021, van Kempen et al¹¹ found 42 studies reporting deep learning-based glioma segmentation models. Currently, only 1 tool has received the US Food and Drug Administration (FDA) approval for this task.¹³

In the context of lesion detection and characterization using MRI, diverse techniques and pulse sequences are available, each possessing distinct inherent attributes that allow the highlighting of particular aspects of different tumors' pathology. The radiologists' expertise relies on using different magnetic resonance sequences to best depict these various aspects of tumors and other diseases. Patients with GBM routinely undergo a standard brain tumor MRI protocol, including fluid-attenuated inversion recovery (FLAIR), T1 weighted image (WI), T2WI, and contrast-enhanced T1-weighted (CET1) MRI.¹⁴ CET1 is a crucial sequence that provides essential insights into tumor location, degree of enhancement and aggressiveness, edema, and necrotic core. The conspicuity of contrast-enhancing lesions (CELs) in relation to the background parenchyma is affected by several factors, including

the ultrastructural features of the blood-brain barrier, magnetic field strength, concentration of gadolinium-based contrast agents, relaxivity properties, time elapsed since injection, and MRI acquisition technique.^{15–18}

Over the past decade, inversion recovery gradient echo (IR-GRE) techniques, such as the 3-dimensional (3D) T1 Magnetization Prepared - RApid Gradient Echo, have emerged as popular pulse sequences for CET1 MRI. These 3D, thin-section techniques can be reformatted into various planes and leverage the inversion recovery preparation pulse to emphasize the contrast between gray and white matter in the brain.^{19,20} IR-GRE techniques such as Magnetization Prepared - RApid Gradient Echo have been recommended for use in brain tumor clinical trials by the Consensus Recommendations for a Standardized Brain Tumor Imaging Protocol in Clinical Trials and the modified Response Assessment in Neuro-Oncology criteria,^{14,21} depicting anatomy very well with excellent signal-to-noise ratio and spatial resolution. However, compared with T1-weighted 3D fast spin echo (FSE) techniques such as CUBE, Sampling Perfection with Application-optimized Contrast using different flip-angle Evolutions (SPACE), and volume isotropic turbo spin echo acquisition (VISTA), IR-GRE images generally yield poorer conspicuity of postgadolinium CELs, particularly if those lesions are within the relatively bright white matter of IR-GRE (which contributes to diminished contrast-to-noise ratio [CNR] of CELs) or if near the enhancing arterial and venous vessels of IR-GRE.^{22–25} This limitation can hinder the accurate detection of lesions, particularly those with subtle contrast enhancement.

Mugler et al²⁶ developed the SPACE sequence as part of their efforts to improve the quality of MRI imaging. SPACE, a 3D-FSE sequence, features variable flip-angle refocusing pulses and can be performed at 3 or 1.5 T. Compared with other techniques, the 3D-FSE method used by SPACE offers greater sensitivity to gadolinium enhancement, particularly in white matter where the CNR of CELs is increased.^{24,25} It can better withstand the negative impact of tumoral hemosiderin and calcium deposits, which can hinder the visualization of enhancement²⁷ (Figure 1). As a "black blood" technique that largely

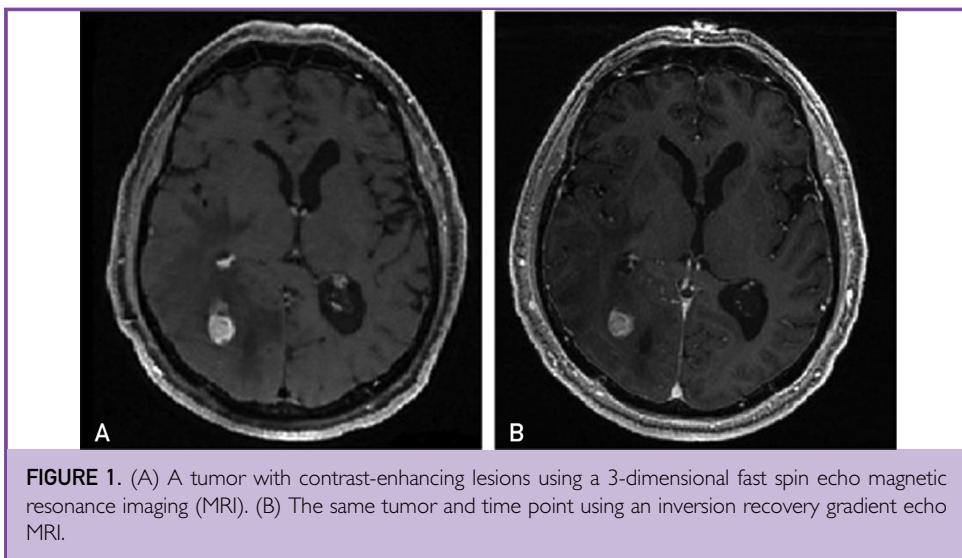


FIGURE 1. (A) A tumor with contrast-enhancing lesions using a 3-dimensional fast spin echo magnetic resonance imaging (MRI). (B) The same tumor and time point using an inversion recovery gradient echo MRI.

suppresses signals related to gadolinium within blood vessels, 3D-FSE improves the conspicuity of small enhancing lesions in the brain periphery near subarachnoid vessels.^{25,28–30}

The successful segmentation of components on MRI is contingent on accurately differentiating their associated signals. To achieve this, it is common practice to use multiple MRI sequences (FLAIR, T1WI, T2WI, and CET1), which provide varying tissue contrasts, for GBM segmentation. Although studies have demonstrated that manual segmentation using 3D-FSE and IR-GRE can result in discrepancy among enhancing lesion components, to our knowledge, no research has assessed how automated segmentation models differentiate between these areas when using IR-GRE or 3D-FSE.¹⁷ Given the escalating utilization and advantages of 3D-FSE, we conducted this study to address this knowledge gap.

PATIENTS AND METHODS

This retrospective study was reviewed and exempted from the requirement for institutional review board approval. We conducted a series of experiments to assess the impact of training and testing segmentation models on CET1 images acquired with IR-GRE or 3D-FSE pulse sequence protocols.

Data Set

We used 3 distinct data sets: the publicly available University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) data set containing multiple pulse sequences and associated segmentation masks for patients with GBM.³¹ This data set comprises 501 cases with T1WI, T2WI, CET1, and FLAIR pulse sequences. The CET1 image sets in this data set are acquired using the IR-GRE protocol.³¹

The other 2 private data sets were collected from the Mayo Clinic institutional imaging database. We selected 100 patients with GBM diagnosed with isocitrate dehydrogenase wild-type tumors after January 1, 2018. We collected each patient's preoperative FLAIR, T1WI, T2WI, and CET1 MRIs. Both data sets contained T1WI, T2WI, and FLAIR images with similar protocols. However, 50 patients had CET1 acquired using the IR-GRE acquisition protocol, and the other 50 patients had CET1 acquired using the 3D-FSE acquisition protocol.

In the IR-GRE acquisition protocol, the TE was observed to range from 1.86 to 23 milliseconds, having a median value of 3.05 milliseconds. The repetition time extended from 6.13 to 2300 milliseconds, with the median being 400 milliseconds. The inversion time ranged from 450 to 100 milliseconds, with a median of 900 milliseconds. The 3D-FSE

acquisition protocol was used, where the TE varied between 8.15 and 44 milliseconds, with a median value of 12.65 milliseconds. The repetition time spanned from 400 to 1959.76 milliseconds, with a median of 600 milliseconds. For both protocols, MRI scanners with magnetic field strengths of both 3 and 1.5 T were used. Both the IR-GRE and J3D-FSE image sets contained scans with slice thicknesses ranging from 3 to 5 mm. We selected image sets with dimensions of 512×512 for the x-axis and y-axis and with slice counts ranging from 24 to 60.

For making the gold-standard segmentation mask for 100 Mayo Clinic image sets, a primary set of masks were computed using a publicly available glioma segmentation model (HD-GLIO; <https://github.com/NeuroAI-HD/HD-GLIO-AUTO>) and then manually reviewed by a neuroradiologist (B.J.E.) with 29 years of experience and a neuroradiology fellow in training from our institution (S.K.R.).^{32,33} The data sets consisted of 3 distinct mask classes, wherein the class denoted “0” referred to the background, “1” denoted edema, and “2” denoted CEL. Because the primary distinction between 3D-FSE and IR-GRE lies in their contrast enhancement, we kept the focus of our segmentation and performance comparison analysis on the CEL class.

We divided data sets into training and test sets. Because the UCSF data set was used for pretraining, it was subsequently split into training and validation sets, with 450 samples allocated to the training set and 51 to the validation set. The 2 private data sets (1 with 50 patients having IR-GRE and 1 with 50 patients having 3D-FSE) were divided into 35-patient training sets and 15-patient test sets. The FLAIR and CET1 pulse sequences were selected for model training and testing from a pool of collected sequences. Using T2 and T1 sequences in the model did not yield improvement in results and impeded the training process.^{34,35} The HD-GLIO-AUTO was used to coregister each patient’s FLAIR and CET1 sequences in the Mayo data set, whereas the UCSF data set had already undergone coregistration and preprocessing.

Model

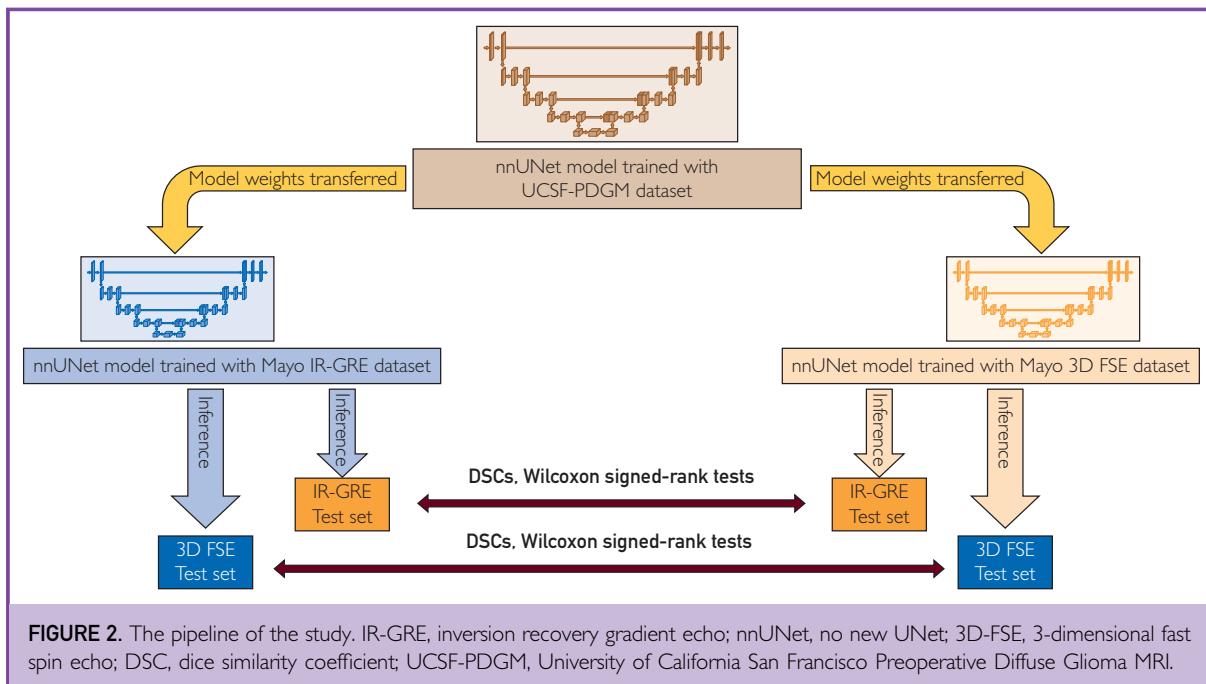
We used no new UNet (nnUNet) to train our models.³⁶ The nnUNet is a deep

learning-based segmentation model that has the capability to customize itself for various tasks in the biomedical field automatically. This includes configuring the preprocessing, network architecture, training, and postprocessing stages. The process relies on a combination of fixed parameters, interdependent rules, and empirical decisions to make key design choices. With nnUNet, there is no need for manual hyperparameter tuning. Presently, nnUNet is the leading-edge tool for segmentation tasks.^{33,37} Fundamentally, a 3D UNet lies at the heart of the nnUNet model, which processes patches that are $128 \times 128 \times 128$ in size. The network comprises an encoder-decoder architecture that connects the 2 pathways through skip connections. The encoder includes 5 levels of convolutional layers with the same resolution and strided convolution downsampling. The decoder follows the same design, with transpose convolution upsampling and convolution applied to concatenated skip features from the encoder branch at the same level. After each convolution operation, a leaky rectified linear unit with a slope of 0.01 and batch normalization was used. The FLAIR and CET1 volumes were concatenated and used as 2-channel input.

Training

We applied nnUNet version 1 for 1000 epochs to each training set. We used the stochastic gradient descent method with an initial learning rate of 0.01 and with a batch size of 2. To enhance generalization, we applied data augmentation during training. This was done on the fly and included techniques such as random rotation and scaling, elastic deformation, gamma scaling, mirroring, and additive brightness augmentation.

The first model was trained and validated on the UCSF-PDGM training set. In the next step, we used the weights of the previous models and fine-tuned the model with Mayo Clinic private data sets. We conducted the fine-tuning process twice: once using the data set containing 35 IR-GRE image sets and once using the data set containing 35 3D-FSE image sets. We used 5-fold crossvalidation for the models, and the best model for each fold was saved. We conducted the experiments with Pytorch 1.13. We used 4



NVIDIA A100 GPUs for our computational tasks. The training took an average of 90 seconds for each epoch.

Performance Comparison

We used the mean dice similarity coefficients (DSCs) as the performance metric for the comparison analysis. DSC is a commonly used metric for evaluating the performance of a segmentation model in medical image analysis, computer vision, and other fields where segmentation is used.³⁸ It measures the similarity between the predicted segmentation mask and the ground truth segmentation mask. The formula for the DSC is as follows:

$$2 \times (\text{intersection}) / (\text{union} + \text{intersection})$$

where intersection refers to the number of pixels classified as positive by the model and ground truth; union refers to the number of pixels classified as positive by the model, ground truth, or both. The dice score ranges from 0 to 1, with 1 indicating perfect agreement between the predicted and ground truth masks.

We also computed the 95th percentile Hausdorff distance (HD95), which is a variant

of the traditional Hausdorff distance. While the Hausdorff distance measures the maximum distance between the boundaries of 2 sets (such as ground truth and predicted masks), HD95 considers only the largest 95% of these distances. This adjustment makes HD95 less sensitive to outliers than the standard Hausdorff distance. Specifically, it calculates the distance within which 95% of all points from one boundary lie from the nearest point on the other boundary. This metric provides a more robust measure of spatial accuracy in segmentation by focusing on most boundary points and offering a realistic view of the precision of object contours. HD95 is particularly useful for assessing segmentation accuracy because it complements metrics like the DSC.

Statistical Analyses

The statistical analysis was conducted using the SPSS software package, version 27.0 (IBM). Normal distribution assumptions were assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. We calculated 2 different DSCs for each case in the 3D-FSE test set. One DSC was calculated using the gold standard and predicted the mask obtained from

TABLE. Dice Scores Obtained From the Inference Process on 2 Distinct Test Sets Using 2 Trained Models

Case No.	3D-FSE test set				Case No.	IR-GRE test set			
	Dice similarity coefficient from the model trained with 3D-FSE images	95th percentile Hausdorff distance	Dice similarity coefficient from the model trained with IR-GRE images	95th percentile Hausdorff distance		Dice similarity coefficient from the model trained with 3D-FSE images	95th percentile Hausdorff distance	Dice similarity coefficient from the model trained with IR-GRE images	95th percentile Hausdorff distance
	3D-FSE-01	0.98657394	2.00	0.92796691	1.00	IR-GRE-01	0.919188422	1.00	0.26451549
3D-FSE-02	0.92114801	1.00	0.866511041	17.49	IR-GRE-02	0.923418145	1.00	0.915960046	1.00
3D-FSE-03	0.593035821	18.27	0.116941342	21.12	IR-GRE-03	0.940072501	1.00	0.964708401	1.00
3D-FSE-04	0.948219096	1.00	0.920096194	1.00	IR-GRE-04	0.920271072	1.73	0.926250265	1.00
3D-FSE-05	0.954590467	1.00	0.918892838	1.00	IR-GRE-05	0.943799187	1.00	0.934584069	1.00
3D-FSE-06	0.953783423	1.00	0.952771946	1.00	IR-GRE-06	0.840494408	38.53	0.92400654	1.00
3D-FSE-07	0.969573031	1.00	0.927997562	1.00	IR-GRE-07	0.937335806	1.00	0.961940193	1.00
3D-FSE-08	0.906564037	1.00	0.936158135	1.00	IR-GRE-08	0.93983324	1.00	0.954972532	1.00
3D-FSE-09	0.959106679	1.41	0.914656837	1.00	IR-GRE-09	0.864130435	1.73	0.884445599	2.00
3D-FSE-10	0.8511646208	9.90	0.704100187	19.74	IR-GRE-10	0.905986022	1.00	0.934689899	1.00
3D-FSE-11	0.98318633	1.00	0.96032081	1.00	IR-GRE-11	0.935465106	1.00	0.927665312	1.00
3D-FSE-12	0.906616185	7.07	0.971174116	2.00	IR-GRE-12	0.953264108	1.00	0.915632347	1.00
3D-FSE-13	NAN ^a	NAN ^a	NAN ^a	NAN ^a	IR-GRE-13	0.836125874	6.78	0.711017647	21.00
3D-FSE-14	0.965717792	1.00	0.907016845	1.00	IR-GRE-14	0.941141344	1.00	0.949803021	1.00
3D-FSE-15	0.952668198	1.00	0.915657094	1.00	IR-GRE-15	0.942733403	1.00	0.965552733	1.00
Average of dice scores	0.925991043		0.869137986		Average of dice scores	0.918736555		0.897179597	
SD	0.107190941		0.276314885		SD	0.041902273		0.197989519	
Median of dice scores	0.953783423		0.918892838		Median of dice scores	0.937335806		0.926250265	

^aThe NAN case lacked any contrast-enhancing component in their tumor region that could be segmented.

IR-GRE, inversion recovery gradient echo; NAN, not a number; 3D-FSE, 3-dimensional fast spin echo.

fine-tuning the model with the 3D-FSE training set, and the other DSC was calculated using the gold standard and predicted the mask obtained from fine-tuning the model with the IR-GRE training set. Similarly, for each case in the IR-GRE test set, we performed the same calculation and obtained 2 DSC values. Subsequently, we conducted the Wilcoxon test to compare the 2 DSCs for each case in the set. Owing to the nonnormal distribution of our data, we conducted a 2-tailed, nonparametric paired Wilcoxon signed-rank test for paired measurements. Figure 2 shows the study pipeline.

RESULTS

Table summarizes the DSCs for the CEL segmentation of the 2 models for each patient in the test set. To clarify, each of the 15 cases in every test set belonged to entirely distinct patients. For instance, IR-GRE-01 represents an entirely separate patient and exhibits a completely different lesion compared with 3D-FSE-01.

The calculated *P* values comparing 2 DSC sets using for cases in the 3D-FSE and IR-GRE test sets was .02 and .61, respectively. Figure 3 presents 2 MRI slices, highlighting a GBM tumor lesion on 3D-FSE and IR-GRE sequences.

DISCUSSION

In this study, we developed 2 distinct segmentation models, each specializing in the delineation of GBM lesions within either IR-GRE or 3D-FSE MRI protocol. Subsequently, we compared these models, evaluating their performance on separate test sets comprising cases obtained from the respective imaging protocols. Our findings reveal that the model specifically trained on the 3D-FSE protocol outperformed its counterpart in segmenting GBM lesions (even in the case of IR-GRE image sets). This superior performance was attributed to the heightened contrast visualization afforded by the 3D-FSE MRI protocol, enhancing the efficacy of lesion segmentation.

We followed the conventional practice of using a 70% to 30% split to partition our

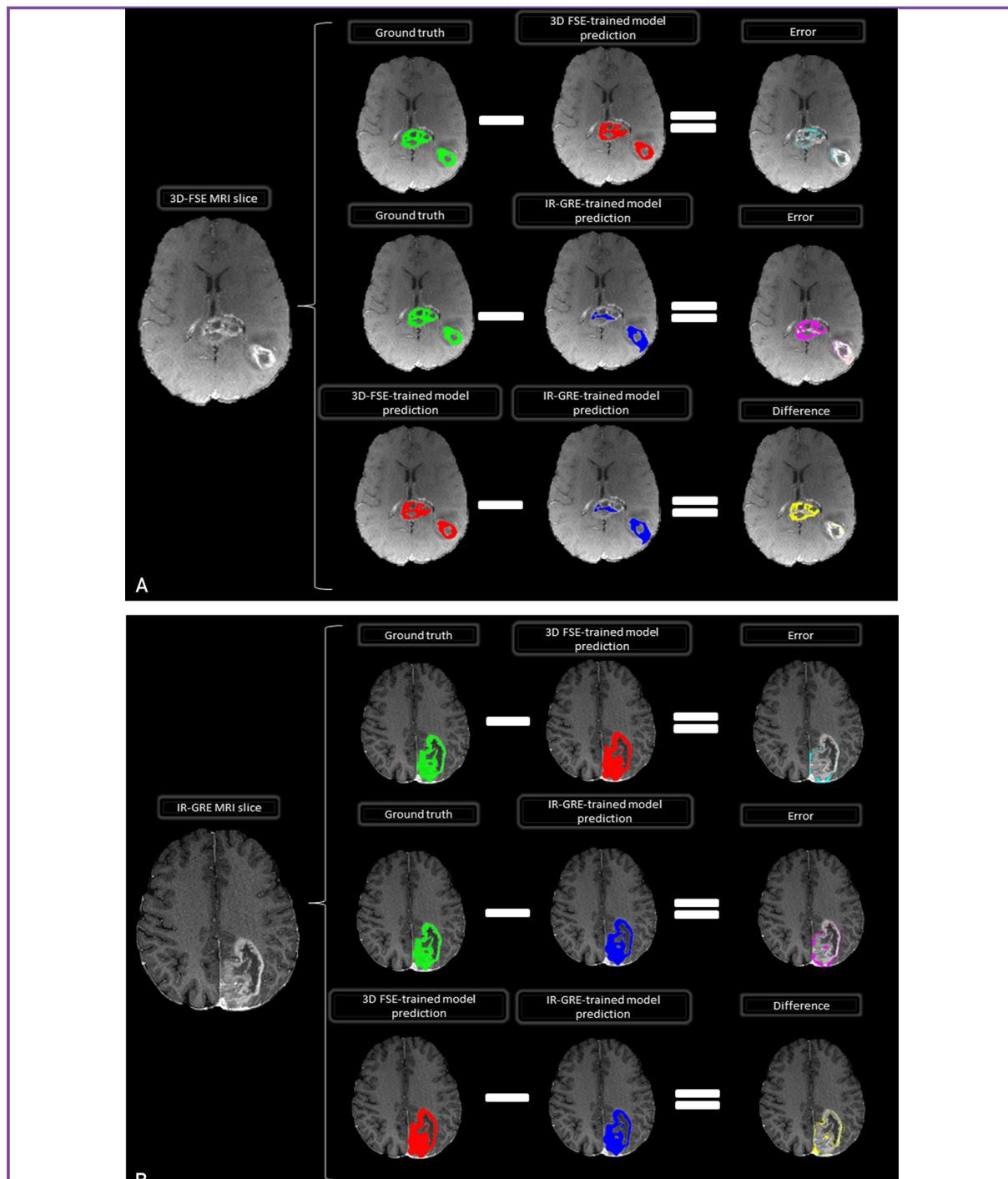


FIGURE 3. Two MRI slices showcasing a glioblastoma tumor lesion. (A) A 3D-FSE MRI slice is displayed, with colors representing the ground truth, predicted, error, and difference masks overlaid on the image. (B) An example of an IR-GRE MRI slice, depicting the ground truth, predicted, error, and difference masks using different colors. IR-GRE, inversion recovery gradient echo; 3D-FSE, 3-dimensional fast spin echo.

image data sets into training and test sets, respectively.³⁹ Given our initial pretraining phase involving 450 cases, it was our supposition that including an additional 35 cases featuring different imaging protocols would sufficiently help the model learn the distinct features of each data set. Moreover, the incorporation of 15 cases for the precise computation of lesion-specific visual attributes diminished the likelihood of encountering stochastic or random outcomes in our study.

Regarding the results in Table, it is important to mention that, in a single instance (3D-FSE-13), the MRI scan did not exhibit any CEL components. Our data selection process deliberately omitted manual inspection to prevent any bias toward cases with more conspicuous lesions. Consequently, such occurrences were accommodated within the data set without bias, and we kept this case. Furthermore, our segmentation models encountered challenges in certain cases, such as 3D-FSE-03, where the lesions proved to be particularly intricate, posing difficulties for both models during the segmentation process.

Several research articles have described the superiority of 3D-FSE image acquisition in identifying small tumors and detecting subtle changes associated with CELs.^{25,28,29} First, Incorporating 3D-FSE sequences into the clinical imaging protocol for patients has a practical implication of improving the detection of regions with a more aggressive grade of malignancy within GBMs. This is crucial for accurately selecting biopsy samples, ensuring correct histologic grading, and making informed treatment decisions. Moreover, maximizing tumor resection has been linked to better patient outcomes, and image-guided approaches have the potential to increase the extent of resection.⁴⁰ Therefore, optimizing the accuracy of target delineation enables precise surgical procedures tailored to individual patients. In the context of planning radiotherapy, the use of low-conspicuity images may hinder the delineation of target volume margins, resulting in suboptimal dose delivery to all regions of active tumor growth and/or unnecessary dose delivery to surrounding critical brain structures.^{41,42} As the utilization of automated segmentation continues to increase in GBM delineation clinical practice and

3D-FSE is becoming a popular sequence, it is crucial to determine whether training with this pulse sequence will impact tumor segmentation model performance or not.¹⁷

To our knowledge, this is the first study that conducted a comparative analysis of the impact of 3D-FSE in automated segmentation models in a general context. According to the findings of our study, models trained on 3D-FSE images demonstrate superior performance in segmenting CEL on both 3D-FSE and IR-GRE image sets. Specifically, we observed an increase of 0.057 in the average of DSCs for the test set cases with 3D-FSE image sets and 0.022 for the test set cases with IR-GRE image sets. We hypothesized that this is due to the superior CNR of the 3D-FSE images, which assists the model in learning the edges and detecting the lesions more effectively. By incorporating 3D-FSE image sets or performing a single step of fine-tuning, GBM segmentation models can be optimized and further refined, thereby improving their performance. Our findings have implications for medical centers with access to 3D-FSE capabilities and for the broader research community using publicly available GBM segmentation models, facilitating improved accuracy and clinical decision-making in the context of GBM segmentation and volumetric analysis.

In this study, we note that our model was pretrained on a data set consisting of 501 IR-GRE cases, which was necessary to achieve an acceptable level of performance for comparative analysis. The models trained solely on 50 cases, without the benefit of pretraining, yielded average DSCs of 0.68, which fell short of the desired threshold for establishing fair comparisons. Ideally, pretraining with the same number of 3D FSE cases would have been preferred for conducting a comprehensive comparative analysis. However, the unavailability of such a data set, coupled with the high costs associated with annotating 501 cases, made this impractical. Nonetheless, although not ideal, our approach of pretraining on 501 IR-GRE cases and subsequently fine-tuning on 50 3D FSE or IR-GRE cases demonstrated the potential benefit of incorporating a fine-tuning step with 3D FSE data.

A limitation of our study arose from our use of segments from 100 patients with GBM, primarily because of the significant time and cost associated with data annotation. Nonetheless, to address the challenge posed by a limited patient count, we initially conducted a pretraining step using a publicly available data set. In our future research efforts, we plan to expand our data set with an increased collection of 3D-FSE images. This will primarily focus on improving both size and quality of the data, thereby enriching the resources available for this study. Another limitation of the concept of adding 3D-FSE to the segmentation model training process is the lack of access to 3D-FSE in all centers. However, as the utilization of this pulse sequence continues to expand, we anticipate including more 3D-FSE CET1 image sets in publicly available data sets. This would enable future studies to incorporate a larger and more diverse range of data, fostering improved understanding and performance of models in the field of medical imaging and analysis. An additional constraint of our study pertains to the generalizability of our 3D FSE model, which primarily applies to the specific acquisition protocols used in the Mayo Clinic.

CONCLUSION

The study findings indicate that models trained on 3D-FSE images exhibit enhanced performance in segmenting CEL across both 3D-FSE and IR-GRE image sets. The optimization and refinement of GBM segmentation models can be achieved incorporating 3D-FSE image sets through a fine-tuning step, leading to improved performance. In future endeavors, we aim to expand our data set with additional 3D-FSE images, focusing on enhancing the size of data available for this research.

POTENTIAL COMPETING INTERESTS

Given his role as Editorial Board Member, Dr Bradley Erickson had no involvement in the peer-review of this article and had no access to information regarding its peer-review.

ACKNOWLEDGMENTS

Drs Moassefi and Faghani contributed equally to this work.

Abbreviations and Acronyms: **FLAIR**, fluid-attenuated inversion recovery; **CET1**, contrast-enhanced T1-weighted; **CEL**, contrast-enhancing lesion; **DSC**, dice similarity coefficient; **HD95**, 95th percentile Hausdorff Distance; **FSE**, fast spin echo; **GBM**, glioblastoma multiforme; **IR-GRE**, inversion recovery gradient echo; **MRI**, magnetic resonance imaging; **nnUNet**, no new UNet; **TE**, echo time; **3D**, 3-dimensional; **UCSF-PDGM**, University of California San Francisco Preoperative Diffuse Glioma MRI

Correspondence: Address to Bradley Erickson, MD, PhD, Mayo Clinic, 200 1st Street, SW, Rochester, MN 55905 (Bje@Mayo.edu; Twitter: @slowvak).

ORCID

Bradley J. Erickson:  <https://orcid.org/0000-0001-7926-6095>

REFERENCES

1. Ostrom QT, Gittleman H, Farah P, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro Oncol*. 2013; 15(suppl_2):ii1-ii6. <https://doi.org/10.1093/neuonc/not151>.
2. Tanimi AF, Juweid M. *Epidemiology and Outcome of Glioblastoma*. Codon Publications; 2017. <https://doi.org/10.15586/codon.glioblastoma.2017.ch8>.
3. Fu Q, Cheng QG, Kong XC, et al. Comparison of contrast-enhanced T1-weighted imaging using DANTE-SPACE, PETRA, and MPRAGE: a clinical evaluation of brain tumors at 3 Tesla. *Quant Imaging Med Surg*. 2022;12(1):592-607. <https://doi.org/10.21037/qims-21-107>.
4. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks. *Med Image Anal*. 2017;35:18-31. <https://doi.org/10.1016/j.media.2016.05.004>.
5. Lim DSW, Makmur A, Zhu L, et al. Improved productivity using deep learning-assisted reporting for lumbar spine MRI. *Radiology*. 2022;305(1):160-166. <https://doi.org/10.1148/radiol.220076>.
6. Moassefi M, Faghani S, Khosravi B, Rouzrokh P, Erickson BJ. Artificial intelligence in radiology: overview of application types, design, and challenges. *Semin Roentgenol*. 2023;58(2):170-177. <https://doi.org/10.1053/j.sro.2023.01.005>.
7. Ermiş E, Jungo A, Poel R, et al. Fully automated brain resection cavity delineation for radiation target volume definition in glioblastoma patients using deep learning. *Radiat Oncol*. 2020;15(1):100. <https://doi.org/10.1186/s13014-020-01553-z>.
8. Juan-Albarraín J, Fuster-García E, Manjón JV, et al. Automated glioblastoma segmentation based on a multiparametric structured unsupervised classification. *PLoS One*. 2015;10(5):e0125143. <https://doi.org/10.1371/journal.pone.0125143>.
9. Porz N, Bauer S, Pica A, et al. Multi-modal glioblastoma segmentation: man versus machine. *PLoS One*. 2014;9(5):e96873. <https://doi.org/10.1371/journal.pone.0096873>.
10. Wu Y, Zhao Z, Wu W, Lin Y, Wang M. Automatic glioma segmentation based on adaptive superpixel. *BMC Med Imaging*. 2019;19(1):73. <https://doi.org/10.1186/s12880-019-0369-6>.
11. van Kempen EJ, Post M, Mannil M, et al. Performance of machine learning algorithms for glioma segmentation of brain MRI: a systematic literature review and meta-analysis. *Eur Radiol*. 2021; 31(12):9638-9653. <https://doi.org/10.1007/s00330-021-08035-0>.
12. Zhang X, Hu Y, Chen W, Huang G, Nie S. 3D brain glioma segmentation in MRI through integrating multiple densely connected 2D convolutional neural networks. *J Zhejiang Univ Sci B*. 2021;22(6):462-475. <https://doi.org/10.1631/jzus.B2000381>.
13. Neosoma, Inc., a Medical Technology Company Focused on Advancing the Treatment of Brain Cancers. Announced Today It Has Received U.S. Food and Drug Administration (FDA) 510(k) Clearance for Its First AI-Based Neuro-

- Oncology Software, Neosoma HGG. Published September 29, 2022. <https://www.businesswire.com/news/home/20220929005284/en/Neosoma-Inc-a-Medical-Technology-Company-Focused-on-Advancing-the-Treatment-of-Brain-Cancers-Announced-Today-It-Has-Received-U.S.-Food-and-Drug-Administration-FDA-510-k-Clearance-for-Its-First-AI-Based-Neuro-Oncology-Software-Neosoma-HGG>. Accessed March 2, 2023.
14. Ellingson BM, Bendszus M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. *Neuro Oncol.* 2015;17(9):1188-1198. <https://doi.org/10.1093/neuonc/nov095>.
 15. Krautmacher C, Willinek WA, Tschampa HJ, et al. Brain tumors: full- and half-dose contrast-enhanced MR imaging at 3.0 T compared with 1.5 T—initial experience. *Radiology.* 2005; 237(3):1014-1019. <https://doi.org/10.1148/radiol.2373041672>.
 16. Cha S, Lupo JM, Chen MH, et al. Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *AJNR Am J Neuroradiol.* 2007;28(6):1078-1084. <https://doi.org/10.3174/ajnr.A0484>.
 17. Daniell L, Riccetti GC, Distefano D, et al. Brain tumor-enhancement visualization and morphometric assessment: a comparison of MPRAGE, SPACE, and VIBE MRI techniques. *AJNR Am J Neuroradiol.* 2019;40(7):1140-1148. <https://doi.org/10.3174/ajnr.A6096>.
 18. Abu Khalaf N, Desjardins A, Vredenburgh JJ, Barboriak DP. Repeatability of automated image segmentation with BraTumIA in patients with recurrent glioblastoma. *AJNR Am J Neuroradiol.* 2021;42(6):1080-1086. <https://doi.org/10.3174/ajnr.A7071>.
 19. Brant-Zawadzki M, Gillan GD, Nitz WR. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence—initial experience in the brain. *Radiology.* 1992;182(3):769-775. <https://doi.org/10.1148/radiology.182.3.1535892>.
 20. Mugler JP III, Brookeman JR. Rapid three-dimensional T1-weighted MR imaging with the MP-RAGE sequence. *J Magn Reson Imaging.* 1991;1(5):561-567. <https://doi.org/10.1002/jmri.1880010509>.
 21. Ellingson BM, Wen PY, Cloughesy TF. Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics.* 2017;14(2):307-320. <https://doi.org/10.1007/s13311-016-0507-6>.
 22. Mugler JP III, Brookeman JR. Theoretical analysis of gadopentate dimeglumine enhancement in T1-weighted imaging of the brain: comparison of two-dimensional spin-echo and three-dimensional gradient-echo sequences. *J Magn Reson Imaging.* 1993;3(5):761-769. <https://doi.org/10.1002/jmri.1880030512>.
 23. Schmitz BL, Aschoff AJ, Hoffmann MHK, Grön G. Advantages and pitfalls in 3T MR brain imaging: a pictorial review. *AJNR Am J Neuroradiol.* 2005;26(9):2229-2237.
 24. Pui MH, Fok EC. MR imaging of the brain: comparison of gradient-echo and spin-echo pulse sequences. *AJR Am J Roentgenol.* 1995;165(4):959-962. <https://doi.org/10.2214/ajr.165.4.7677001>.
 25. Suh CH, Jung SC, Kim KW, Pyo J. The detectability of brain metastases using contrast-enhanced spin-echo or gradient-echo images: a systematic review and meta-analysis. *J Neurooncol.* 2016; 129(2):363-371. <https://doi.org/10.1007/s11060-016-2185-y>.
 26. Mugler JP III, Bao S, Mulkern RV, et al. Optimized single-slab three-dimensional spin-echo MR imaging of the brain. *Radiology.* 2000;216(3):891-899. <https://doi.org/10.1148/radiology.216.3.r00au46891>.
 27. Tins B, Cassar-Pullicino V, Haddaway M, Nachtrab U. Three-dimensional sampling perfection with application-optimised contrasts using a different flip angle evolutions sequence for routine imaging of the spine: preliminary experience. *Br J Radiol.* 2012;85(1016):e480-e489. <https://doi.org/10.1259/bjr/25760339>.
 28. Komada T, Naganawa S, Ogawa H, et al. Contrast-enhanced MR imaging of metastatic brain tumor at 3 tesla: utility of T1-weighted SPACE compared with 2D spin echo and 3D gradient echo sequence. *Magn Reson Med Sci.* 2008;7(1):13-21. <https://doi.org/10.2463/mmrs.7.1.3>.
 29. Kato Y, Higano S, Tamura H, et al. Usefulness of contrast-enhanced T1-weighted sampling perfection with application-optimized contrasts by using different flip angle evolutions in detection of small brain metastasis at 3T MR imaging: comparison with magnetization-prepared rapid acquisition of gradient echo imaging. *AJNR Am J Neuroradiol.* 2009;30(5):923-929. <https://doi.org/10.3174/ajnr.A1506>.
 30. Kaufmann TJ, Smits M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. *Neuro Oncol.* 2020;22(6): 757-772. <https://doi.org/10.1093/neuonc/noaa030>.
 31. Calabrese E, Villanueva-Meyer JE, Rudie JD, et al. The University of California San Francisco preoperative diffuse glioma MRI dataset. *Radiol Artif Intell.* 2022;4(6):e220058. <https://doi.org/10.1148/rai.220058>.
 32. Kickingereder P, Isensee F, Tursunova I, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol.* 2019;20(5):728-740. [https://doi.org/10.1016/S1470-2045\(19\)30098-1](https://doi.org/10.1016/S1470-2045(19)30098-1).
 33. Isensee F, Jäger PF, Kohl SAA, Petersen J, Maier-Hein KH. Automated design of deep learning methods for biomedical image segmentation. *arXiv [cs.CV].* Published online April 17, 2019, <http://arxiv.org/abs/1904.08128>. <https://doi.org/10.1038/s41592-020-01008-z>.
 34. Holtzman Gazit M, Faran R, Stepovoy K, Peles O, Shamir RR. Post-operative glioblastoma multiforme segmentation with uncertainty estimation. *Front Hum Neurosci.* 2022;16:932441. <https://doi.org/10.3389/fnhum.2022.932441>.
 35. Fyllingen EH, Stensjøen AL, Berntsen EM, Solheim O, Reinertsen I. Glioblastoma segmentation: comparison of three different software packages. *PLoS One.* 2016;11(10):e0164891. <https://doi.org/10.1371/journal.pone.0164891>.
 36. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnUNet: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;18(2):203-211. <https://doi.org/10.1038/s41592-020-01008-z>.
 37. Zhang Y, Liu S, Li C, Wang J. Rethinking the dice loss for deep learning lesion segmentation in medical images. *J Shanghai Jiaotong Univ.* 2021;26(1):93-102. <https://doi.org/10.1007/s12204-021-2264-x>.
 38. Faghani S, Khosravi B, Zhang K, et al. Mitigating bias in radiology machine learning: 3. Performance metrics. *Radiol Artif Intell.* 2022;4(5):e220061. <https://doi.org/10.1148/rai.220061>.
 39. Rouzrok P, Khosravi B, Faghani S, et al. Mitigating bias in radiology machine learning: I. Data handling. *Radiol Artif Intell.* 2022; 4(5):e210290. <https://doi.org/10.1148/rai.210290>.
 40. Sanai N, Berger MS. Surgical oncology for gliomas: the state of the art. *Nat Rev Clin Oncol.* 2018;15(2):112-125. <https://doi.org/10.1038/nrclinonc.2017.171>.
 41. Zhao F, Li M, Kong L, Zhang G, Yu J. Delineation of radiation therapy target volumes for patients with postoperative glioblastoma: a review. *Onc Targets Ther.* 2016;9:3197-3204. <https://doi.org/10.2147/OTT.S104241>.
 42. Farace P, Giri MG, Meliàdò G, et al. Clinical target volume delineation in glioblastomas: pre-operative versus post-operative/pre-radiotherapy MRI. *Br J Radiol.* 2011;84(999):271-278. <https://doi.org/10.1259/bjr/10315979>.